A consensus approach to sentiment analysis

Orestes Appel[1,2], Francisco Chiclana[1], Jenny Carter[1], and Hamido Fujita[3]

[1] De Montfort University, Centre for Computational Intelligence (CCI), Leicester, United Kingdom. orestes.appel@my365.dmu.ac.uk,{chiclana,jennyc}@dmu.ac.uk
[2] Mount Royal University (Bissett School of Business), Calgary, Alberta, Canada. oappel@mtroyal.ca
[3] Iwate Prefectural University (IPU), Takizawa, Iwate, Japan. issam@iwate-pu.ac.jp

Abstract. There are many situations where the opinion of the majority of participants is critical. The scenarios could be multiple, like a number of doctors finding commonality on the diagnosing of an illness or parliament members looking for consensus on a specific law being passed. In this article we present a method that utilises Induced Ordered Weighted Averaging (IOWA) operators to aggregate a majority opinion from a number of Sentiment Analysis (SA) classification systems, where the latter occupy the role usually taken by human decision-makers. Previously determined sentence intensity polarity by different SA classification methods are used as input to a specific IOWA operator. During the experimental phase, the use of the IOWA operator coupled with the linguistic quantifier 'most' (IOWA$_{most}$) proved to yield superior results compared to those achieved when utilising other techniques commonly applied when some sort of averaging is needed, such as arithmetic mean or median techniques.

Keywords: Hybrid Sentiment Analysis Method; Naïve Bayes; Maximum Entropy; Consensus; Majority Support; Sentiment Aggregation; IOWA operator.

1 Introduction

Group decision making (GDM) is a task where a number of agents get involved in a decision process to generate a value that represents their individual decisions in the group process [4]. The *arithmetic mean* and the *median* are central tendency values widely used in homogeneous GDM where experts are equally important [9]. In heterogeneous GDM, Yager's Induced Ordered Weighted Averaging (IOWA) operator [10] is widely used because it allows for different importance degrees to be implemented but also because it allow for the implementation of the concept of 'majority'. In the Sentiment Analysis (SA) context, any of the possible SA classification methods available can be considered an agent. The aim is to obtain a collective classification value that reflects *the opinion of the majority* of the SA classification methods. Experiments have been conducted using the IOWA operator coupled with the linguistic quantifier 'most' (IOWA$_{most}$) to implement the concept of majority with three SA classification methods: Naïve Bayes [6], Maximum Entropy [2], and the Hybrid Approach to the SA problem devised in [1].

The remainder of this paper is organised as follows: Section 2 covers the basic concept of IOWA operator and the derivation of its associated weighting vector. Section 3 addresses the role of IOWA operators in the achieving of fuzzy majority in collective decision-making. In order to provide context, Section 4 covers the hybrid method introduced by the authors in [1], as the approach presented in this article represents an enhancement to this method in terms of obtaining a majority sentiment classification opinion. Section 5 covers the experimental results obtained when applying the proposed majority based methodology, and section 6 closes the paper with some conclusions and a brief discussion of possible further work.

2 IOWA Operators

Yager's IOWA operator [10] has been proved to be extremely useful in group decision making problems because it allows to implement the concept of *fuzzy majority* [11].

Definition 1 (IOWA Operator). *An IOWA operator of dimension n is a mapping $IOWA\colon (\mathbb{R} \times \mathbb{R})^n \longrightarrow \mathbb{R}$, which has an associated set of weights $W = (w_1, \cdots, w_n)$ to it, verifying $w_i \in [0,\ 1]$, $\sum_{i=1}^{n} w_i = 1$, such that*

$$IOWA\left(\langle u_1, a_1\rangle, \ldots, \langle u_n, a_n\rangle\right) = \sum_{i=1}^{n} w_i \cdot a_{\sigma(i)}, \tag{1}$$

and $\sigma\colon \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ is a permutation function such that $u_{\sigma(i)} \geq u_{\sigma(i+1)}$, $\forall i = 1, \ldots, n-1$.

In the above definition the reordering of the set of values to aggregate, $\{a_1, \ldots, a_n\}$, is induced by the reordering of the set of values $\{u_1, \ldots, u_n\}$ associated to them. In [11], Yager proposed the following approach to obtain the IOWA associated weighting vector: Let $Q : [0,1] \to [0,1]$ be a function such that $Q(0) = 0$, $Q(1) = 1$, and $Q(x) \geq Q(y)$ for $x > y$ corresponding to a fuzzy set representation of a proportional monotone quantifier. Then,

$$w_i = Q\left(i/n\right) - Q\left((i+1)/n\right) \tag{2}$$

Some examples of linguistic quantifiers are *"at least half"*, *"most of"* and *"as many as possible"*, which have been proposed [11] to be represented by using function:

$$Q(r) = \begin{cases} 0 & \text{if } 0 \leq r < a \\ \frac{r-a}{b-a} & \text{if } a \leq r \leq b \\ 1 & \text{if } b < r \leq 1 \end{cases} \tag{3}$$

with the values $(0, 0.5)$, $(0.3, 0.8)$ and $(0.5, 1)$ for (a, b), respectively [5].

3 IOWA based Fuzzy Majority in GDM

In [7], Pasi and Yager elaborate that one of the possible semantics of IOWA operators is that of being drivers of a *majority opinion*. What is required is an operator that computes an average-like aggregation of the "**majority** of values that are similar". In [7], the authors establish that "similar values must have close positions in the induced ordering in order to appropriately be aggregated". Hence, the final output of an IOWA operator should reflect the opinion of the majority *if* similar values are closer to each other in the induced vector. Then, what is required is the ability to calculate the similarities between the values to be aggregated in order to compute "the values of the inducing variable of the IOWA operator" [7]. In order to support this, a binary support function, Sup, is introduced, where $Sup_\alpha(a,b)$ expresses the support from b for a at an α level of desired tolerance based on the concept that "the more similar two values are, the more they support each other":

$$Sup_\alpha(a_i, a_j) = \begin{cases} 1 & \text{if} \quad |a_i - a_j| < \alpha \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The higher the tolerance is, the less it is imposed that the two values *have to be closer to each other to support each other*. If we were to aggregate a set of values and we wanted to order them in increasing order of support, then for each value the sum of its support values is computed with respect to the rest of the values to be aggregated [7]. These overall supports are utilised as the values of the order inducing variable. Thus, the use of an adequate support function will enable the induction of an ordering based on proximity, which is key to understanding how IOWA operators generate a *majority-based aggregation* of the values to be aggregated via the linguistic quantifier *most* (as presented in Eq. (3) with values $(a,b) = (0.3, 0.8)$). Also, Pasi and Yager's strategy implies that the construction of the weighting vector appropriately implements more *influence* in the aggregation result from the most supported individual values. Consequently the following process for the construction of the weighting vector from the induced support values is proposed:

1. Include in the definition of the overall support for a_i the similarity of the value a_i with itself:
$$t_i = s_i + 1. \tag{5}$$
2. On the basis of the t_i values, the weights of the weighting vector are computed as follow:
$$w_i = Q\left(\frac{t_i}{n}\right) \Big/ \sum_{j=1}^{n} Q\left(\frac{t_j}{n}\right) \tag{6}$$
"The value $Q(t_i/n)$ denotes the degree to which a given member of the considered set of values represents the *majority*" when the linguistic quantifier Q is used.

4 A Hybrid Approach to the SA Problem at the Sentence Level aimed to opinion consensus

In [1], we describe a hybrid model for the SA (HSC/HAC) problem at the sentence level that is based on semantic rules, smart NLP techniques and fuzzy sets. The IOWA approach for aggregation presented in this article will be used to complement the aforementioned HSC/HAC model with the aim to arrive at a consensus sentiment classification opinion in SA [3] representing the opinion of the majority of approaches available to address the SA problem, as depicted in Fig. 1. Formally, the problem to address is how to determine the subjectivity intensity polarity for a given sentence S_k using the outputs of several classification systems. In a way, each method to be used and applied to the aforementioned sentence S_k, can be seen as an 'agent/person' giving her opinion on whether the sentence S_k is positive or negative. In our context, we would like to *aggregate the polarity intensity* value of sentence S_k measured by using different classification methods. Hence, the final polarity value will be the 'induced aggregation of the majority' of the subjectivity intensity polarity of sentence S_k when one takes into consideration all the different contributions of all the participating classification methods. The different applied classification methods will issue their individual *opinions*, just as individual agents use their own judgement.

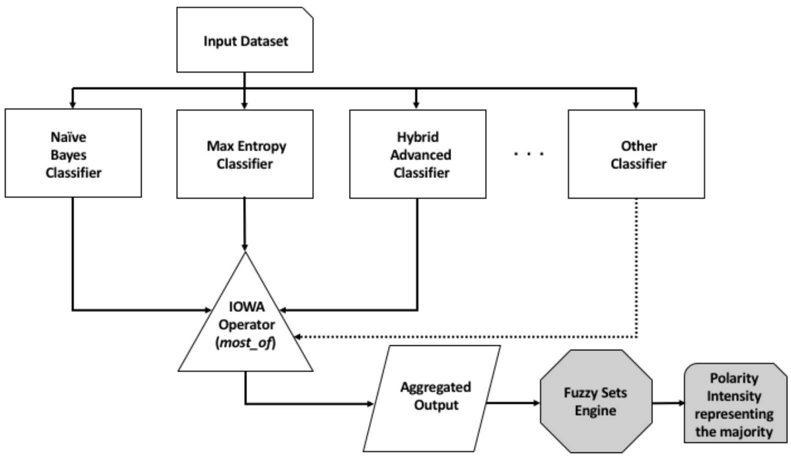

Fig. 1. *$IOWA_{most}$ operator aggregating classifier methods outputs*

5 Experimental results obtained applying $IOWA^{\alpha}_{most}$ aggregation

The $IOWA^{\alpha}_{most}$ operator performance was evaluated against both the *Arithmetic mean* and the *Median* performances (see Table 1). Experiments have been performed for both the Movie Review Dataset (`http://www.cs.cornell.edu/people/pabo/movie-review-data/`) and the Twitter Dataset *Sentiment140* (`http://help.sentiment140.com/for-students`). In order to use the output of all classifiers as an input to the $IOWA^{\alpha}_{most}$ process all participating scores have been converted to the interval $[0,1]$, where S_k corresponds to any sentence in the test dataset and $\{m_1, m_2, \ldots, m_n\}$ represents the different classification methods being aggregated ($n \geq 2$), then:

$$IOWA^{\alpha}_{most}[S_k](m_1, m_2, \ldots, m_n) = \Theta^{S_k} \tag{7}$$

Once the aggregation with the semantic representing the *opinion of the majority* has been computed, then Eq. (7) corresponds to the intensity level in which the value Θ belongs. The aggregated value Θ^{S_k} will take on the value x in $\mu_{\tilde{A}}(x)$ and in consequence a proper linguistic label belonging to set $G = \{Poor,\ Slight,\ Moderate,\ Very,\ Most\}$ will be generated to represent the polarity intensity (how positive or how negative) of a given sentence S_k [1]. The datasets used in the proof of concept count each (positive occurrences and negative occurrences) with 5,331 sentences. We have annotated 500 sentences, approximately 10%, assigning each of them a value $v_k \in G$. These were estimated by looking at the classification outcomes of the three classifiers we are utilising as inputs and estimating a linguistic label in G that is representative of the opinion of the majority. However, before we applied the IOWA$_{most}$ operator, we combined directly the results of the three chosen methods using the *Arithmetic mean* and the *Median*. The outcomes, which are summarised in Table 1 below, are not as good as those obtained by using the IOWA operator. This fact, shows that the IOWA operator does a much better job at aggregating the individual outcomes of the three aforementioned techniques, by giving more weight to the leaning opinion of the majority. In essence, by properly weighting the advise of the three methods (NB, ME and the HSC/HAC approach) we do obtain a more realistic aggregation effect that represents the thinking of the majority. The main difference between the results obtained when using different tolerance values (0.3 and 0.5) when IOWA$^{\alpha}_{most}$ is applied, is not in whether the outcome will distance itself from representing the opinion of the majority, but rather to which *linguistic label* (Poor, Slight, Moderate, etc.) a specific sentence will be assigned. Depending on the majority value calculated a sentence classified as 'Moderate' with a tolerance of 0.3 could now be labelled as 'Very' in terms of intensity, when the tolerance value changes to 0.5. In reality, the lower the *tolerance*, the more demanding the IOWA operator is on how closely the values in the aggregation support each other (see Table 1 for the experiments results).

6 Conclusions and further work

IOWA operators can certainly play a significant role in aggregating the opinions of a number of sentiment classification systems. The aforementioned operator works by producing a value that gets significantly closer to the collective opinion of the participants. The IOWA$^{\alpha}_{most}$ used in this article conveys the semantic of

Semantic	Median	Mean	$\text{IOWA}_{most}^{\alpha=0.3,\ 0.5}$
Represents opinion of the majority	337	388	500
Does **not** represent opinion of the majority	163	112	0
% of success	67.40	77.60	100

Table 1. All aggregation methods compared

the opinion of the majority ('most'). Its performance in identifying the *intensity of the opinion of the majority*, according to our experiments, surpassed the one exhibited by *Arithmetic Mean* and *Median* techniques. In essence, $\text{IOWA}_{most}^{\alpha}$ produces a larger pull towards the values that support each other, driving the results in the direction of what the majority reflects. In terms of further work, we believe there are some avenues that could be pursued in the short-term: (a) Investigate other OWA operators that could potentially produce a better aggregation representing the semantic *majority opinion* and (b) Utilise the OWA measure of *dispersion*, which calculates the degree to which all aggregates are used in the resulting final aggregation [8].

References

1. O. Appel, F. Chiclana, J. Carter, and H. Fujita. A Hybrid Approach to the Sentiment Analysis Problem at the Sentence Level. *Knowledge-Based Systems*, 2016, doi: 10.1016/j.knosys.2016.05.040.
2. C. M. Bishop. *Pattern Recognition and Machine Learning.* Springer Science + Business Media, LLC, 1st edition, 2006.
3. F. J. Cabrerizo, F. Chiclana, R. Al-Hmouz, A. Morfeq, A. S. Balamash, and E. Herrera-Viedma. Fuzzy decision making and consensus: challenges. *Journal of Intelligent & Fuzzy Systems*, 29(3):1109–1118, 2015.
4. F. Herrera and E. Herrera-Viedma. Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, 115:67–82, 2000.
5. J. Kacprzyk. Group decision making with a fuzzy linguistic majority. *Fuzzy Sets and Systems*, 18:105–118, 1986.
6. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the Association for Computational Linguistics (ACL-02) Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 10:79–86, 2002.
7. G. Pasi and R. R. Yager. Modeling the concept of majority opinion in group decision making. *Inf. Sci.*, 176(4):390–414, Feb 2006.
8. J. Peláez, J. M. Doña, and J. Gómez-Ruiz. Analysis of OWA operators in decision making for modelling the majority concept. *Applied Mathematics and Computation*, 186:1263–1275, 2007.
9. J. Perkins. *Python Text Processing with NLTK 2.0 Cookbook.* Packt Publishing, 2010.
10. R. Yager and D. Filev. Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 29(2):141–150, Apr 1999.
11. R. R. Yager. Quantifier guided aggregation using OWA operators. *International Journal of Intelligent Systems*, 11(1):49–73, 1996.